

AD-A124 447

SOFTWARE TOOLS FOR VOICE RECOGNITION RESEARCH(U)
VIRGINIA POLYTECHNIC INST AND STATE UNIV BLACKSBURG
COMPUTER S... J M HOERNEMANN ET AL. NOV 82 CS18-82-11

1/1

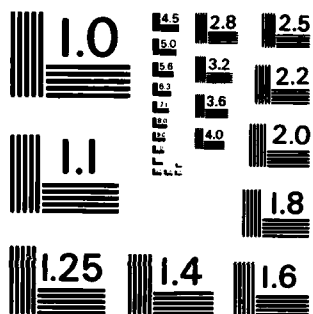
UNCLASSIFIED

NO0014-B1-K-0143

F/D 0/2

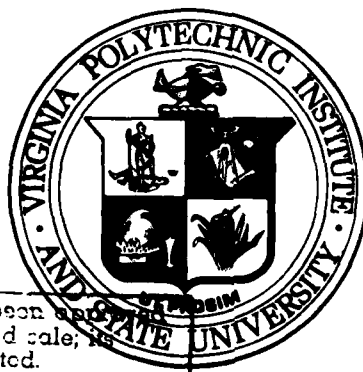
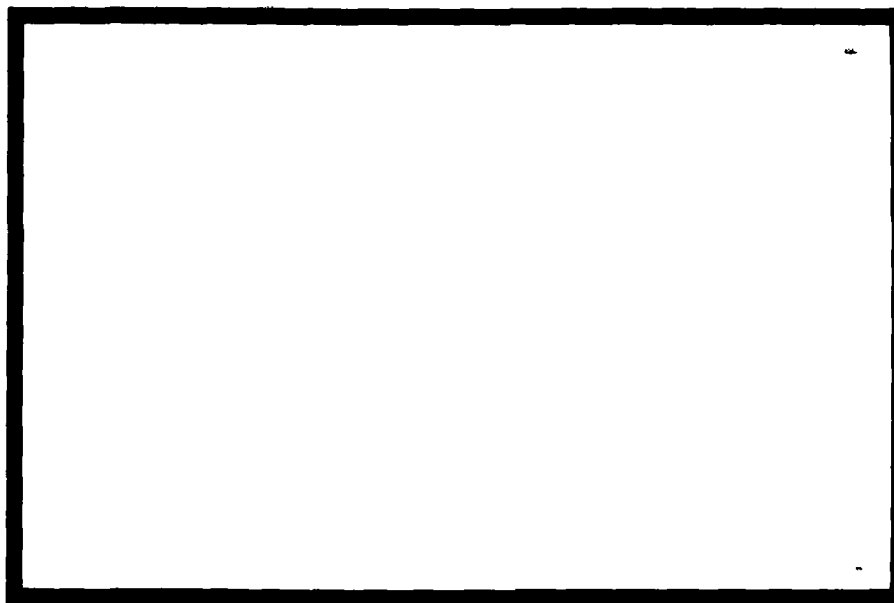
NL

END
DATE
FILMED
3 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ADA 124447



This document has been approved for public release and sale; its distribution is unlimited.

SP-100

Virginia Polytechnic Institute and State University

Computer Science

Industrial Engineering and Operations Research

BLACKSBURG, VIRGINIA 24061

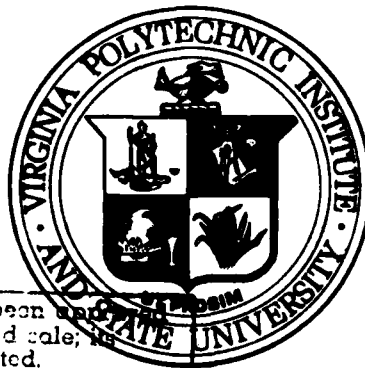
83 02 016 011

DTIC FILE COPY

ADA 124447

SOFTWARE TOOLS FOR VOICE RECOGNITION
RESEARCH

John W. Hoernemann
Joseph F. Maynard
Beverly H. Williges



This document has been approved for public release and sale; its distribution is unlimited.

DTIC
S

Virginia Polytechnic Institute
and State University

Computer Science

Industrial Engineering and Operations Research

BLACKSBURG, VIRGINIA 24061

83 02 016 011

DTIC FILE COPY

SOFTWARE TOOLS FOR VOICE RECOGNITION
RESEARCH

John W. Hoernemann
Joseph F. Maynard
Beverly H. Williges

TECHNICAL REPORT

Prepared for
Engineering Psychology Group, Office of Naval Research
ONR Contract Number N00014-81-K-0143
Work Unit Number SRO-101

Approved for Public Release; Distribution Unlimited

Reproduction in whole or in part is permitted
for any purpose of the United States Government

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CSIE-82-11	2. GOVT ACCESSION NO. <i>AD-A224</i>	3. RECIPIENT'S CATALOG NUMBER <i>447</i>
4. TITLE (and Subtitle) SOFTWARE TOOLS FOR VOICE-RECOGNITION RESEARCH	5. TYPE OF REPORT & PERIOD COVERED Technical Report	
7. AUTHOR(s) John W. Hoernemann, Joseph F. Maynard, & Beverly H. Williges	6. CONTRACT OR GRANT NUMBER(s) N00014-81-K-0143	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science/Industrial Engineering Virginia Polytechnic Institute & State University Blacksburg, VA 24061	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N(42); RR04209; RK0420901; NRSRO-101	
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Code 442 800 North Quincy Street Arlington, VA 22217	12. REPORT DATE November 1982	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES 82	
	15. SECURITY CLASS. (of this report) Unclassified	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) voice recognition, voice input, human factors, software tools, input devices		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report describes software tools provided to interface an Interstate Electronics Corporation Voice Recognition Module (VRM) to a VAX 11/780 computer. The purpose of the tools is to provide a flexible and easy to use set of software modules for experimenters to implement VRM host functions in various task environments. Software to provide analysis of recognition data is also presented. These statistics aid experimenters in developing vocabularies for various implementations of voice input. The voice input hardware and software tools are being used to explore the use of voice as a component of human/computer dialogues.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102- LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research under ONR Contract Number N00014-81-K-0143, Work Unit Number NR SRO-101. The effort was supported by the Engineering Psychology Group, Office of Naval Research, under the technical direction of Dr. John J. O'Hare. Mr. Gabe Moschetti and Dr. Sam Viglione, Interstate Electronics Corporation, provided technical data on the recognition algorithm.



Accession For	
NTIS	<input checked="checked" type="checkbox"/>
DTIC	<input type="checkbox"/>
INPEC	<input type="checkbox"/>
Dist	
A	

TABLE OF CONTENTS

INTRODUCTION	1
VOICE INPUT SYSTEM HARDWARE	3
Voice Terminal (VOTERM)	3
Voice Recognition Module (VRM)	8
Microphone	9
Modem	9
Cables	9
Host	10
Auxiliary terminal	10
Connecting the Hardware	10
Independent operation	10
Slaved terminal	10
PARAMETER MANIPULATION	12
Recognition Algorithm	12
Word Boundaries (T1 and T2)	14
End-of-word Counter (ETHL)	16
Minimum Word Length Counter (MINSM)	17
Reject Threshold (RTHL)	17
Difference-Score Criterion	18
REFERENCE PATTERN DEVELOPMENT	20
VRM Hardware Functions	20
Train	20
Update	21
Recognize	21
Upload	21
Download	22
Software Tools	22
Command format	24
Data files	26
Modifying the software	26
Software Procedures for Using Hardware Functions	26
SETCHARS.PRO	26
SETFLAGS.PRO	28
SETPARMS.PRO	28
SETREJECT.PRO	28
TRAIN.PRO	28
UPDATE.PRO	29
RECOGNIZE.PRO	29
UPLOAD.PRO	30
DOWNLOAD.PRO	30
Software Support Procedures	31
GET_INPUT.PRO	31
YES_or_NO.PRO	32
ERROR.PRO	32
DELAY.FOR	32

Experimenter Interface	32
Creating vocabularies	32
Using the software	32
Parameters	33
Set reject threshold	33
Train	33
Update	35
Recognize	36
File-word patterns	37
Get-word patterns	38
Recognition Problems	38
TASK ENVIRONMENTS	41
Software Modifications Required	41
Feedback and Error-Correction Alternatives	43
Feedback	43
Error correction	44
Experimental Task Environments	46
Prompt/recognition	46
Maze	47
Form-filling	47
GENIE	49
DATA ANALYSIS	50
Software	51
Command Functions	56
Alphabetic summary	56
Confusion problems	57
Delta-score table	58
Exit	58
Information loss	58
List delta < specified value	58
Misrecognitions	58
More data	58
Prompted vs. recognized Table	59
Statistics	59
Recognized for RTHL and DELTA	59
REFERENCES	61
APPENDIX A: VRM Command Summary	62
APPENDIX B: Sample Output from Data Analysis Routines	68
DISTRIBUTION LIST	83

INTRODUCTION

The Voice Input System (VIS) consisting of the VOTERM (VOice TERMinal) chassis and power supply and the VRM (Voice Recognition Module) circuit board, manufactured by Interstate Electronics Corporation, is being used to evaluate voice input in human-computer dialogues. This system provides speaker-dependent, discrete-word recognition of up to 100 words or phrases at a relatively low cost. However, before experimental dialogues could be developed, the VIS had to be interfaced with a VAX 11/780 computer. This report describes both the physical (hardware) and software interfaces and includes descriptions of:

- (1) the voice recognition hardware,
- (2) the Interstate Electronics recognition algorithm,
- (3) the selection of recognizer parameters,
- (4) the software procedures for developing reference patterns,
- (5) the task environments developed that allow voice input, and
- (6) automatic procedures for analyzing voice recognition data.

The VIS will be used to explore the use of voice as a component of human/computer dialogues. This report is designed as an experimenter's guide for configuring the hardware and using the software tools written for the VIS. Along with the descriptions of the hardware and software are brief explanations of potential problems and how to avoid them.

The primary purpose of the VIS software tools is to provide procedures that are easy to use and to modify for interfacing the voice recognition equipment with a VAX computer. Early in the software-development phase it became apparent that the software would be used by a wide range of people

for many different experiments. In addition, there were very few dialogue guidelines for voice input available to aid in answering specific design-related questions. As a result, flexibility was designed into the VIS software to allow experimenters to change the program's input and output easily. The goal of providing for rapid modification of the input and output of any software is quite often the antithesis of providing a well-designed user interface. However, whenever possible, human factors principles were followed in the design of the computer-experimenter dialogue in addition to maintaining the flexibility necessary to design a variety of research studies. The five major goals in the design of these software tools for voice recognition research were to:

- (1) provide procedures to implement the hardware functions of the VIS,
- (2) provide procedures that are easy to modify;
- (3) provide procedures that are easy to use;
- (4) provide procedures that are fully protected against errors generated either by users or by the hardware; and
- (5) provide procedures that allow voice input to various task environments including GENIE. (See Lindquist, Fainter, Guy, Hakkinen, and Maynard, 1982, for a complete description of the GENIE task environment.

VOICE INPUT SYSTEM HARDWARE

The VIS hardware consists of the VOTERM chassis and power supply and the VRM 102 circuit board which is housed in the VOTERM. Other necessary hardware include a microphone, connecting cables, a terminal, a host computer (VAX 11/780 running VMS Version 2.5 or higher), and, optionally, a modem. Most of the VIS software was written for use with a VT100 terminal.

Voice Terminal (VOTERM)

The VOTERM is a self-contained chassis to house the VRM board and includes a power supply, fuse (1/2 amp), audio indicator LED, Vu meter, audio amplifier, audio-level switch, and fan.

The rear panel of the VOTERM has four DB-25S connectors, two for serial communications to interface with the host and an auxiliary device and two for parallel communication to the host and configuration control. Serial host communication is used in the Human Factors Laboratory for the VIS because the distance to the host does not permit parallel interfacing. Therefore, the connectors (ports) labelled "host" (J2) and "auxiliary" (J4) are the only ones used. When a terminal is slaved to the VIS, both the J2 (Host) and J4 (Aux) ports are used. When the VIS is used as a separate device, only the J2 (Host) port is used. The other two ports (J1 and J3) are never used in a serial configuration. For complete details on the pin functions and numbers for these connectors see the Voice Recognition Module Reference Manual (1981) provided by Interstate Electronics.

The front panel of the VOTERM includes a power switch, a microphone input jack, an audio-level switch, an audio meter, an LED indicator, and the

VRM manual-reset switch (red button). The audio-level switch with settings from 1 to 5 may be adjusted to permit different gain levels depending upon background noise, speaker intensity, and microphone distance. Interstate Electronics recommends a setting of "3" under normal conditions. The audio meter provides user feedback of the audio level. Interstate Electronics personnel suggest that having the meter needle in the red area does not indicate input distortion. The LED processing light indicates when the VRM is processing input (on) or idle (off). It should only be on when audio input is being received. If the audio light always remains on, the VRM board may be improperly connected. This may be remedied by reseating the board. If this is necessary, the software procedures must be restarted because power to the board has probably been interrupted. If reseating the board does not correct the problem, it is probably a hardware failure and the board may have to be repaired or replaced. The red switch permits the user to reset the VRM manually without host intervention. Upon reset, all six flags are set to zero (see Table 1), all hardware settings are read (see Table 2), and parameters are set at default values (see Table 3). Upon completion of the reset the VRM sends the host a signal. Whenever power is applied to the VIS, the mode flags, hardware switches, and default parameter values are also read and/or reset, and the same signal is sent from the VRM as for a reset. CAUTION: DO NOT PRESS THE RESET BUTTON. As the result of a hardware reset or power-up, the VRM returns to the default framing characters. For VAX compatibility these framing characters, used in all software written for the VIS, differ from the default characters, and a software procedure must be called to change these characters.

TABLE 1

VRM Mode Flags

<u>Flag</u>	<u>Purpose</u>	<u>VIS Software</u>
1	Provide extended data during recognition	1
2	Acknowledge all host commands	1
3	Provide high-speed parallel upload and download of reference patterns	0
4	Not used	0
5	Transmit all host characters from host to Port 2	1
6	When utterance is too long (>250 significant samples), output LL in place of vocabulary item number.	1

TABLE 2

VRM Hardware Switches with Settings to Connect to the VAX
(Serial Communication)

Switch SA - Baud Rate

SA-1	ON	9600
SA-2	OFF	4800
SA-3	OFF	2400
SA-4	OFF	1200
SA-5	OFF	600
SA-6	OFF	300
SA-7	OFF	150
SA-8	OFF	110

Switch SB - Word Format

SB-1	ON	Auto Port 2 Transfer
SB-2	OFF	Not used
SB-3	OFF	1 stop bit
SB-4	ON	No parity
SB-5	OFF	8 bit words
SB-6	OFF	Not used
SB-7	OFF	Not used
SB-8	Off	Not used

Switch SC - Multipurpose

SC-1	ON	8 megahertz
SC-2	OFF	Not used
SC-3	OFF	Parallel handshaking
SC-4	OFF	Parallel handshaking
SC-5	ON	Serial I/O
SC-6	OFF	'R' sent as reset acknowledgement
SC-7	ON	CR used as terminator
SC-8	OFF	Echo mode

Switch SD - Current Loop or RS-232

SD-1	C1	RS-232
SD-2	C3	RS-232

Switch SE - preamplifier and Port 2 logic level

Out	bypass microphone preamplifier
Port 2	RS-232 logic levels for Port 2

TABLE 3

User-Selectable Parameters for the Interstate Electronics Voice Input System

<u>Parameter</u>	<u>Definition</u>	<u>Purpose</u>	<u>Range</u>	<u>Default</u>
T1	Threshold value for initial significant sample	Detect onset of speech; ignore background noise	16-64	32
T2	Threshold value for subsequent significant samples	Detect continuation of speech; determine maximum length of spoken word	8-64	16
End of Word Threshold (ETHL)	Maximum number of non-significant samples during utterance	Detect end of speech; ignore brief silence during word	8-64	32
Minimum Samples (MINSM)	Minimum number of significant samples for an utterance	Determine minimum word length; reject abrupt noises	16-32	16
Reject Threshold (RTHL)	Minimum delta score for recognition	Precision of match in recognition	0; 98-128	None
Difference Score	Minimum difference between delta scores of top two words	Reduce confusion between top scoring words	0-128	None

Voice Recognition Module (VRM)

The VRM is a single circuit board consisting of a CPU, 4k bytes of ROM to store the processing algorithms, 4k bytes of RAM to store reference patterns, analog circuitry for spectrum analysis of speech, and communications circuitry and internal switches for parallel or serial operation. The useful audio bandwidth is approximately 260 to 6000 Hz. To provide a high quality signal an equalizer boosts the signal level in the frequency bands above 750 Hz. VRM hardware functions for training, recognition, and communication will be described fully in the sections dealing with the VAX software implementation.

There are five different sets of internal switches on the VRM board to control the following functions:

SA: Serial Baud Rate

SB: Serial Word Format

SC: Multipurpose

- serial or parallel communication
- termination character
- echo options
- parallel handshaking mode
- power on/reset acknowledgement

SD: Current loop or RS-232

SE: On-board amplification

Correct settings for these switches to connect the VIS to the VAX using serial communication are given in Table 2. Several of these switches work in tandem, such as SB-3, SB-4, and SB-5, and are set in relation to each other. For the VAX, the combination yielding 1 stop bit, no parity, and 8-bit words is used. It is unlikely that any of these switch settings would

need to be changed unless the VIS were connected to a communication line at a different baud rate (switch SA).

Microphone

A Shure Brothers SM-10A headband-mounted microphone is provided for use with the VIS. Other high quality, noise-cancelling microphones can be used. However, the user is cautioned that consistent microphone positioning is critical for optimal recognition performance (see Nye, 1982a; 1982b). The SM-10A should be positioned close to the lower lip and slightly off to the side. The foam windscreen should not be removed because it protects against wind noise and explosive breath-sounds.

The microphone plugs into the front panel of the VOTERM and can be turned on or off by the user with an inline switch. When the microphone is not in use, it should be turned off. For experimental purposes the microphone is also connected to a tape recorder so that user utterances can be directly recorded.

Modem

The modem (modulator/demodulator) in use at the present time is a Develcon product. A modem that can be operated at least at 9600 baud should be used to connect the VIS to the VAX. Using a 9600-baud line minimizes the time required for filing and loading reference patterns serially.

Cables

A special cable must be used to connect the VIS to the modem because the VIS itself is configured like a modem. This causes a problem when it is connected to the VAX through a modem because the pin information does not match using a standard cable.

Host. The cable marked "VOTERM use only" is a "reverse cable" and was made to solve the problem of interfacing the VOTERM with the modem. This cable was made by reversing the transmit and receive lines (pins 2 and 3) and connecting the Request To Send (RTS) and Clear To Send (CTS) lines (pins 4 and 5) together (shorting them). This interchange of pin functions reverses the cable so that the VOTERM is treated as a terminal by the VAX.

Auxiliary terminal. Connecting (slaving) a terminal to the VIS can cause additional problems. With a VT100 terminal, any standard cable may be used. However, with an HP terminal, a special cable is necessary. This cable needs to have pins 11 and 23 canceled or disconnected by unsoldering the connections.

Connecting the Hardware

The VIS may be connected in two different configurations, either on a separate line to the VAX or with a slaved terminal. All software written for the VIS used in the Human Factors Laboratory uses separate lines for the VIS and the terminal. However, both configurations are described.

Independent operation. Connect the VIS to the host using a 9600-baud modem and the J2: Host port on the rear panel of the VOTERM. Use only the special cable provided for this purpose (see section on cables). The microphone is plugged into the connector on the front panel of the VOTERM, and the power switch is turned on. Any terminal connected to the VAX may now be used to activate the VIS software.

Slaved terminal. This configuration differs from independent operation in that the terminal is directly connected to the VIS using the J4: Auxiliary port on the rear panel of the VOTERM. All information from the terminal is passed through the VIS prior to being sent to the VAX. This configuration

negates the terminal type-ahead buffer normally available because input from the keyboard may be read as input from the VIS.

PARAMETER MANIPULATION

The VRM allows the user to manipulate five parameters that affect recognizer accuracy. These parameters determine the data samples to be included in the input utterance pattern as well as the stringency of the recognition procedure. To understand the use of these parameters, it is necessary to understand the procedures used by the VRM for data collection and recognition.

Recognition Algorithm

Basically, recognition is a process of counting matching bits between incoming speech and stored reference patterns. Incoming speech is analyzed by a 16-channel comb-filter spectrum-analyzer and is converted to digital form, significant samples are saved and are reduced to a fixed-size pattern, and finally the patterns from incoming utterances are compared to a set of stored reference patterns.

More specifically every 5 msec the output from each of the 16 band-pass filters of the spectrum analyzer is digitized and examined. A basic concept in the VRM recognition algorithm is the comparison of each sample to the last significant sample to determine word boundaries and redundant spectral content (typical of voiced sounds). If the difference between the new sample and the last significant sample (increase in spectral energy) exceeds a threshold value (T_1), the sample is considered to be significant and the onset of speech is detected. Once the beginning of a word is detected, spectral differences in subsequent samples are compared to a second threshold value (T_2); those that exceed T_2 are classified as significant and are saved. If the spectral difference for the sample does not exceed T_2 , the counter for

nonsignificant samples is incremented in order to provide data necessary to detect the end of the utterance. For each significant sample, amplitude normalization is achieved by comparing the frequency energy in each adjacent filter and assigning a 0 or a 1 (depending upon whether the spectral slope is increasing or decreasing between filters) to yield a 15-bit frequency pattern. Data collection continues until a maximum of 250 samples of data are collected or the end of the word is detected. Time normalization of utterances is accomplished by dividing the input buffer of significant data into eight time-intervals and averaging each interval by taking a majority vote on the slope coding. This yields a 120-bit (8 time slots x 15 bits) reference pattern for the utterance regardless of the absolute length of the utterance.

Detecting the end of a word is defined by the counter for nonsignificant samples exceeding a threshold value (ETHL). For the utterance to be processed as a word, the number of significant samples collected must exceed a minimum value (MINSM). This is done to avoid the problem of attempting to recognize abrupt noises.

During training to develop reference patterns, each repetition of the same utterance is logically "anded" to previous utterances. A measure of consistency of the utterances for a specific vocabulary item during training is the number of bits in agreement (NBA). To determine whether a word has been recognized, the incoming 120-bit pattern is compared to each of the stored patterns using a scoring procedure called a delta score. The formula to calculate the delta score between any stored pattern and the current utterance is:

$$\text{Delta Score (I,J)} = 128 * \frac{\text{NBM(I,J)} - 8}{\text{NBA(J)}} \quad (1)$$

where $NBM(I,J)$ is the number of matching bits comparing a stored template to the incoming utterance and $NBA(J)$ the number of bits in agreement during training for the stored reference pattern. The maximum possible delta score is 128.

With the current implementation of the software, two criteria must be met before a word is recognized. First, the hardware requires that the delta score of one or more words must meet or exceed a reject threshold (RTHL), and, second, the software requires that the difference between the delta scores of the top two choices must exceed a difference-score criterion established by the user. If both criteria are met, the reference pattern with the highest delta score is classified as the recognized word. If either criteria is not achieved, a word-reject or not-sure is signalled. This procedure is an expansion of the built-in VRM recognition function which checks only RTHL. Checking for the difference score has been provided through software. The software permits the experimenter to select both the RTHL and the difference-score criterion.

The experimenter may also select values for the threshold for the onset of speech ($T1$), the threshold for continuing speech ($T2$), the minimum number of significant samples required or minimum word-length (MINSM), and the end-of-word threshold or silence-duration permitted (ETHL). These parameters are described more fully in the following sections and are summarized in Table 3.

Word Boundaries ($T1$ and $T2$)

The VRM recognition algorithm determines the beginning of an utterance by detecting an increase in spectral energy relative to that of the silence

interval (includes ambient noise) following the previous utterance. Two parameters establish threshold values for determining which incoming samples are significant and will be stored. For each sample the current spectral energy in the filters is compared to that of the last significant sample. If the difference exceeds the threshold value for the onset of speech (T1), the sample is stored. Once the word is begun, the difference between subsequent samples and the last significant sample is compared to the threshold value for continuing speech (T2) to determine significant samples for storage. Whenever the incoming sample does not differ significantly from the previous sample, it is ignored. In this way redundant information can be discarded.

The selectable range for the speech-onset threshold (T1) is 16-64 with a default setting of 32. T1 must be set small enough so that even the weakest consonants (f, h, m, n) are detected. However, if background noise is substantial, T1 should be increased from the default value. T1 should always be set higher than T2 since the onset of speech is an abrupt change.

For T2 the selectable range is 8-64 with a default setting of 16. The setting of T2 indirectly affects the maximum length of an utterance. The input buffer permits no more than 250 significant samples. With a sampling rate of 5 msec and all incoming samples significant, the longest allowable utterance would be 1.25 sec (5 msec/sample x 250 samples). However, if the value of T2 is increased, more incoming samples will be classified as nonsignificant, and the absolute duration of the longest allowable utterance will be increased. If T2 is too large, it is possible to collect insufficient data (less than MINSM) for recognition. If the vocabulary contains only short words, T2 should be set low to reduce the probability that any incoming

sample will be discarded. However, if T2 is too small, it is possible that the number of nonsignificant samples necessary to detect the end of an utterance (ETHL) will not be satisfied except in a totally quiet environment with minimal breath noise. The values selected for T1 and T2 depend to a great extent upon the composition of the vocabulary.

End-of-Word Counter (ETHL)

Because the VIS is a discrete-word recognizer, a pause is necessary between each word to define the end of the word. Usually a system designer wishes to minimize the length of the required pause. However, many words such as "eight" and "delete" have an internal pause that must be processed without detecting the end of the word prematurely. In addition, a vocabulary might include polysyllabic words or phrases that have internal pauses. Therefore, a tradeoff is necessary to establish a minimum between-word pause for a specific vocabulary such that the end of a word is not detected prematurely and maximum throughput is achieved.

Word/phrase boundaries in the VIS are detected based upon the relative change in spectral energy in the VRM filters from sample to sample. A counter stores the number of nonsignificant samples in which the spectral energy difference between two samples does not exceed the threshold value defined by T2. When the value in this counter exceeds ETHL, the end of speech is detected. The default value for ETHL is 32 which is equivalent to a minimum between-word pause of 160 msec (32 samples x 5 msec/sample). The value for ETHL is user selectable in the range 8 to 64. For rapid speech with brief inter-word pauses, ETHL should be set to a low value. However, when vocabulary words with internal pauses, polysyllabic words, or phrases are included in the vocabulary, ETHL may need to be increased.

The two parameters that determine recognizer response time are the value selected for ETHL (pause length) and the size of the vocabulary which determines the number of reference patterns that need to be compared (processing time).

Minimum Word Length Counter (MINSM)

The VIS has the capability to process discrete words or phrases whose duration is in the range of 80 msec to 1250 msec or more. The minimum word-duration is determined by the parameter MINSM which establishes a criterion for the minimum number of significant samples. The purpose of this parameter is to reduce the probability that the recognizer will try to process abrupt noises as words. When MINSM is set at its default value of 16, the minimum word-length, assuming all incoming samples are significant, would be 80 msec ($5 \text{ msec/sample} \times 16 \text{ significant samples}$). The selectable range for MINSM is 16 to 32.

Reject Threshold (RTHL)

The parameter used during the bit-matching phase of recognition is reject threshold (RTHL). RTHL establishes a level of precision required of the match between incoming patterns and stored patterns as defined by the delta scores calculated using Equation 1. The threshold is given in terms of number of bits in agreement, and although theoretically the range should be 0-128, the actual user-selectable range is 98-128 and 0. If 0 is selected, no input utterances will be rejected. Selecting a high value for RTHL will cause the VIS to reject more invalid utterances and noise but may also result in the rejection of valid utterances.

The delta score for a vocabulary word tends to increase with training. According to Interstate Electronics, the score maximizes between 7 and 10 training passes to a value of 118 to 124. They suggest the use of the following values for RTHL based upon the number of training passes:

<u>Training Passes</u>	<u>RTHL</u>
3-5	100-106
7-10	110-118

In addition, Interstate Electronics provide the following formula to estimate RTHL:

$$\text{RTHL} = 96 + 2 * \text{NTP} \quad (2)$$

where NTP equals the number of training passes. The voice recognition software tools discussed in subsequent sections calculate this value automatically.

Difference-Score Criterion

Through software a second parameter may be used in the bit-matching phase of recognition, the difference score. The difference score establishes a level of certainty or confidence level that the first-choice word is correct and distinctly different from the second-choice word. The difference score defines the minimal separation in delta scores required between the top two scoring words for a recognition to occur. The value is user selectable from 0 to 99. Interstate Electronics suggests that a difference greater than or equal to 6 indicates a high probability (99%) that the words are not being confused. The capability to establish a value for the difference score and check for it

during recognition has been provided by software and is not a built-in function of the VRM.

REFERENCE PATTERN DEVELOPMENT

Hardware Functions

Reference patterns must be developed for every vocabulary item and for the speaker. These patterns are the composite of a number of training patterns through the vocabulary. The VRM provides programmable functions to facilitate the development and storage of reference patterns. These include update, recognize, upload, and download. In addition, RTHL, T1, T2, INSM, the flags, and the control characters can be altered by built-in functions.

The information transmitted between the VAX and the VIS is in terms of ASCII characters. All commands to and responses from the VRM are framed by control characters. However, some of the VRM default control characters are interpreted either by the VAX or the VT100 terminal (when slaved to the VAX). This necessitates changing the control characters. The new control characters were chosen from the set of ASCII characters that is not used in the representation of reference patterns. This was done to ensure that the pattern data would not be interpreted as control characters. Data bits 6, 7, and 8 in reference patterns are always 0011 yielding the ASCII characters 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, :, ;, <, =, >, ?, - during uploading and downloading. The basic VRM functions used to develop reference patterns are described below.

Input. Any contiguous set of words in the vocabulary or a single word may be selected, and the number of training passes may be specified. The VRM utilizes the temporary storage area for reference patterns, provides indices for each input, and generates a reference pattern for each word or phrase.

Update. The update function differs from training only in that the storage area for reference patterns is not initialized; that is, the previous patterns are retained and incorporated with the update passes.

Recognize. The primary built-in function of the VRM is recognition in which incoming utterances are compared with stored patterns. The experimenter may select any subset of the reference patterns to be compared to user utterances during recognition. The vocabulary items selected need not be contiguous. Using this feature of vocabulary subsetting or windowing, one reduces the size of the active vocabulary used in the comparison thereby increasing the probability of correct recognition. Therefore, recognition accuracy and response time should be improved. The recognition command can return one of two sets of data: (1) the winning word and difference between the delta scores of the winning and runner-up words or (2) the winning word, the runner-up word, the difference between the delta scores of the winning and runner-up words, and the winner's score. To receive the extended list of data from recognition, mode flag 1 must be set (see Table 1). All software written for the VIS in the Human Factors Laboratory returns the extended list of information.

Upload. The upload function transmits reference patterns to the host for storage. Assuming no host delays, the approximate time for data transfer with serial operation can be calculated as follows:

$$\text{Transfer Time (sec)} = \frac{(\text{NP}) 680}{\text{Baud Rate}} \quad (3)$$

where NP equals the number of patterns to be transmitted. The last four ASCII characters of each reference pattern consist of two pairs of numbers which provide the number of bits in agreement during training (NBA) and the number of training passes (NTP).

Download. This function transfers patterns from host storage to the VRM replacing all or part of the contents of the VRM RAM. If only part of the 100-word capacity is replaced by downloading new patterns, some of the old reference patterns will remain available in the VRM RAM. The approximate transfer time for downloading is the same as for uploading, assuming no host delays.

Software Tools

A software package of nine procedures incorporating a menu-driven user interface is available to facilitate the use of the built-in VRM functions. Two of the nine procedures used to control the VRM functions are not seen by the experimenter but are used to control how the VRM functions. These two procedures set the control characters (SETCHARS.PRO) used in communication between the VAX and the VRM and set the flags (SETFLAGS.PRO) that control what functions to VRM will perform. These two procedures are run automatically whenever any of the software for the VIS is used. Two additional procedures are used to set the reject-threshold (RTHL) and other parameters (T1, T2, ETHL, and MINSM). The remaining group of five procedures involves the actual operations or functions of the VRM. These five functions are: training patterns, updating patterns, recognizing utterances, filing patterns in host storage, and retrieving patterns from host storage. The name of each software procedure is related to its menu name and function (see Table 4).

TABLE 4

Menu Name, Function, and Procedure Name of Experimenter's
Software Package for Developing and Storing Reference Patterns

<u>Menu</u>	<u>Functions</u>	<u>Procedure Name</u>
Parameters	Sets parameters	SETPARMS.PRO
Set reject-threshold	Sets reject-threshold	SETREJECT.PRO
Get-word-patterns	Retrieves patterns from host	DOWNLOAD.PRO
Train	Develops new patterns	TRAIN.PRO
Recognize	Recognizes words	RECOGNIZE.PRO
Update	Updates old patterns	UPDATE.PRO
File-word-patterns	Stores patterns on host	UPLOAD.PRO
_____	Changes control chars	SETCHARS.PRO
_____	Changes flags	SETFLAGS.PRO

In addition, a set of four procedures have been written to facilitate the user interface. These procedures read and process user input, display error information, and delay erasure of information displayed. Each of these software procedures is described in the following sections.

Incorporated in each of these software tools are three characteristics:

- (1) All input is checked for being of the proper type. If alphabetic characters are expected as input, alphabetic characters will be accepted; any other type of input (integers, real numbers, etc.) will not be accepted, and the user will be asked for the correct input.
- (2) All input is checked for being within the proper bounds.
- (3) Any unexpected response from the VRM will cause an error message. The error message will tell the user exactly what occurred in order to aid in finding the condition that caused the error. The usual cause of error is pushing the reset button on the VOTERM.

Command format. The software procedures used to implement the VRM functions all follow the same general format. Basically, they receive input from the user, format the input, send it to the VRM, receive input from the VRM, format that input and display it for the user. User input and VRM response are always checked for errors. Communication commands and responses from the VRM are always framed in control characters. It should be noted that the default control characters have been replaced with control characters compatible with the VAX (see SETCHARS.PRO). Sending and receiving commands is accomplished through the local variables:

Sending

Host_to_VRM_control_char
Host_to_VRM_control_number
Host_to_VRM_control_letter
VRM_Command

Receiving

VRM_to_Host_control_char
VRM_to_Host_control_number
VRM_to_Host_control_letter
VRM_response

The commands and responses from the VRM (see Appendix A) are PASCAL one-dimensional packed arrays of characters of size 15. The first character of this array is always the VRM_to_Host_control_char, and the second element of this array is either the VRM_to_Host_control_number or VRM_to_Host_control_letter (depending on which command is currently being used). The rest of the array, except the last character read, is information (such as word number) particular to the specific function being used. The last character received (not necessarily element 15 of the array) is the terminator (carriage return).

Appendix A defines the symbols used in the VRM commands and details what is sent to the VRM and what the VRM sends back in response for each command. It should be noted that for readability, the commands in Appendix A have commas inserted between characters (!,1,xx,yy,bb,T). However, in actual use these commands (and the VRM's responses) are all characters strings containing no commas or other delimiters.

Data files. Data required or generated by the software are stored in two types of files. Pattern files contain the voice patterns created by the TRAIN procedure. In addition to some identifying records, these files contain 1 to 100 records of 68 characters each to define the reference patterns in the vocabulary. These files are created using the UPLOAD procedure and are read by the DOWNLOAD procedure. The second type of data file contains character strings of the vocabulary being used. This file is read by the TRAIN, UPDATE, and RECOGNIZE procedures. Each of these files is described more fully in the following section.

Modifying the software. All procedures are fully documented with comments as well as meaningful variable names. To change any of the software, copy all of the procedures to your own directory. In addition, change the "%include" commands in the main program (voice.pas) to include the modified files in your directory.

Software Procedures for Using Hardware Functions

SETCHARS.PRO. SETCHARS redefines the control characters to be used. This redefinition was necessary because the VAX and the VT100 trap the default control characters of the VRM. The new control characters were selected from the set of ASCII characters not used to represent reference patterns (see Reference Pattern Development). Otherwise, confusion would result when downloading reference patterns because the VRM would interpret a reference pattern (or part of one) as a command. Table 5 gives a one-to-one mapping of the default and assigned control characters. SETCHARS is called at the beginning of the voice program, following a reset, and after any error is detected. The user has no control over when SETCHARS is called. The use of the control characters is evident from Appendix A.

TABLE 5

VRM Control Characters: Default and Assigned Values

<u>Purpose</u>	<u>Default Value</u>	<u>Assigned Value</u>
Reset Character	STX	STX
Framing Character 1	DC1	!
Framing Character 2	DC2	&
Framing Character 3	DC3	-
Framing Character 4	DC4	@
Acknowledged	ACK	%
Nonacknowledged	NAK	#
Command Terminator	CR	CR

SETFLAGS.PRO. SETFLAGS sets the six internal flags of the VRM (see Table 1). Each of these flags can be set either on or off, but there is only one proper setting. Therefore, SETFLAGS is called at the beginning of the main program (voice.pas), following a reset, and anytime an error is detected. The user has no control over when SETFLAGS is called.

SETPARMS.PRO. This procedure prompts the user for new values for T1, T2, ETHL, and MINSM. However, if a user determines a set of parameters that works well and will always be used, SETPARMS can be modified to set these parameters automatically when the voice program is first run so that no prompting is necessary.

SETREJECT.PRO. This procedure prompts the user for the reject-threshold value (98-128 or 0). This procedure also calculates a suggested reject-threshold value by using Equation 2. The VRM will only acknowledge reject-threshold values of zero or values in the range from 98 to 128. As with SETPARMS, if a standard reject-threshold value is to be used, the procedure can be modified and called automatically when the voice program is first run.

TRAIN.PRO. TRAIN develops a pattern or template for each vocabulary item. The user is first prompted to enter the numbers of the first and last words to be trained and the number of training passes (0-64). If "0" is entered for the number of training passes, the VRM (not the software) defaults to five training passes if only one word is being trained. If more than one word is being trained, the VRM will use the number of training passes used for the previous set of words trained. If no previous set of words has been trained, the VRM requires a number of training passes greater than zero. Then the software prompts the user to enter a number greater than zero.

TRAIN prompts the user to say each vocabulary item by displaying the word prompt on the screen. This is done by the use of the array "vocabulary" that can contain a maximum of 100 vocabulary items of up to 15 characters each. (The maximum length of each vocabulary item may be increased from 15 by minor changes to the software.) The array is loaded by reading the vocabulary file (see section on data files).

UPDATE.PRO. The UPDATE software is basically the same as the TRAIN procedure. The only differences are that the word "train" has been replaced by "update" everywhere it appeared in the software, and the send and receive control characters are different. However, the VRM hardware function UPDATE differs from the hardware function TRAIN. TRAIN develops new word patterns (old patterns are destroyed), whereas UPDATE refines existing patterns. If a word is trained again versus updated, the existing word pattern will be lost, and a new one will be developed.

RECOGNIZE.PRO. This procedure provides random prompting for one complete pass through the vocabulary. RECOGNIZE displays the word recognized by the VRM or the appropriate error message. RECOGNIZE differs from other VRM functions in that there is no documented method by which to get the VRM out of recognition mode. Recognition is terminated after all words have been prompted once by a software reset to the VRM. A user may abort the recognition mode prior to a complete pass through the vocabulary by entering a CNTL-C on the keyboard. Because the reset changes control characters, flags, and the reject-threshold (undocumented in the VRM manual), the procedures SETCHARS, SETFLAGS, and SETREJECT are called before control is returned to the main program.

Two VRM functions related to RECOGNIZE are not implemented in the training software. The first of these is the extended vocabulary. Using extended vocabulary allows the user to instruct the VRM to recognize multiple subsets of the vocabulary as opposed to the entire vocabulary or a contiguous subset of the vocabulary. It was decided that this function was not required for pattern training. The second feature that is not implemented in RECOGNIZE is "common vocabulary." Common vocabulary increases the size of the active vocabulary by attaching a new set of words to the current active vocabulary. However, there is no difference between this command and reissuing the RECOGNIZE command with a larger vocabulary.

UPLOAD.PRO. UPLOAD creates VAX files of reference patterns. The user is prompted for the name of the file where the reference patterns and other relevant information are to be stored. Anytime an old file is used, UPLOAD will create a new version of the old file with a higher version number. Thus, the old patterns will not be lost.

DOWNLOAD.PRO. DOWNLOAD readies the VRM to receive a VAX file of reference patterns. The user is prompted for the name of the file where these patterns are stored and for the number of patterns to be downloaded. Any contiguous set of patterns can be downloaded. After the file name is read, the first line of the file (the number of reference patterns) is read and the user is presented with this information. The user is asked for the first word number and the last word number to be downloaded (they must designate either the entire file or a contiguous subset of the file). The remainder of the information in the file header is then read. This information is:

Name of subject
File creation date
T1 used in training
T2 used in training
ETHL used in training
MINSM used in training
Reject threshold used in training
Vocabulary file (prompt words) used in training
Number of training passes
Number of update passes

The program then downloads the reference patterns, compares the current settings of the VRM to those in the file, and displays the information for the user. This software actually works by repeating the download command X times (X being the number of patterns requested by the user). The VRM counts characters as they are received (68 characters/word pattern). If the incorrect number of characters is received, DOWNLOAD must be run again.

Software Support Procedures

The software for training and recognition of voice patterns uses four support procedures: GET_INPUT, YES_OR_NO, ERROR, and DELAY.

GET INPUT.PRO. This procedure is used to read all keyboard input. GET_INPUT treats all input as characters and checks for characters in the 0 to 9 range. Characters other than 0 to 9 (including blanks) are not allowed, and if they are detected the user is prompted for the correct input. All proper input is converted to integer before being returned to the calling procedure.

YES OR NO.PRO. This procedure is used to read user input involving the answer to yes/no questions. Users can enter the words "yes" or "no" or the first letters "y" or "n" in either upper or lower case.

ERROR.PRO. This procedure is called if the response from the VRM was other than expected. The procedure displays what was received from the VRM. No analysis of the information is done. Before control is returned to the calling procedure, SETCHARS, SETFLAGS, and SETREJECT are all called.

DELAY.FOR. DELAY is a Fortran procedure that calls system service routines to place the process into hibernation for a desired time period. This is used to provide the user enough time to read displayed information. DELAY can be used to delay any aspect of the program for any desired time period.

Experimenter Interface

Creating vocabularies. A file of vocabulary items to be used to prompt the user during training should be created. During training the words will be prompted in the order in which they appear in the vocabulary file. Each vocabulary item (maximum of 40 characters) is entered on a separate line in capital letters. The vocabulary list can consist of any words or phrases and should be based upon the requirements of the particular experiment or application. Because of the memory capacity of the VRM, vocabulary size is limited to 100 vocabulary items. For convenience in locating vocabularies, these files should be named with a file type of "voc." Any file name (maximum of 9 characters) may be used. For example, the vocabulary to run the maze is named [onr.voice]maze.voc.

Using the software. After insuring that the VIS hardware is correctly connected and power is on, the user should log on to a VAX account using

any VT100 terminal. The training/recognition software resides in the directory [onr.voice]. To use it one types "run [onr.voice]voice." The user will be prompted to enter the file name of the vocabulary to be used during training, such as [onr.voice]maze.voc. A menu (see Table 6) will appear, and the user will be prompted to enter a single character associated with the desired procedure. After entering the appropriate single character and hitting the carriage return, the desired procedure will run. The user is prompted for all necessary data. Whenever a procedure is completed, the menu will again be presented. When these software procedures are used to develop reference patterns for experimental subjects, only the experimenter sees the menu. A brief discussion of each procedure as it appears to the experimenter follows.

Parameters. The set-parameters option prompts the experimenter to enter the desired values for T1, T2, ETHL, and MINSM. For each parameter the current value and selectable range are provided, and the experimenter is asked whether a change in the parameter is desired. If the answer is "yes", a new value for the parameter may be entered.

Set reject-threshold. The set-reject-threshold option sets a criterion for the minimum number of bits in agreement between utterance and reference patterns for word recognition. The current value is provided, and the experimenter is asked whether a change is desired. If so, the number of training passes used to develop the patterns is entered so that the software can calculate a suggested reject-threshold using Equation 2. However, the experimenter may select 0 or any value between 98 and 128.

Train. The train option readies the recognizer to develop reference patterns by initializing the RAM. The user is prompted to say words from a

TABLE 6

Main Menu of Software to Develop Reference Patterns

VOTERM FUNCTIONS

(P)arameters

(S)et reject-threshold

(G)et word patterns

(T)rain

(R)ecognize

(U)pdate

(F)ile word patterns

(E)nd

Enter the desired function letter. _____

vocabulary list stored as a VAX file (see Creating Vocabularies). Word patterns are developed using data from a number of training passes. A training pass usually consists of one enunciation of each word in the vocabulary. However, users should note that the VRM automatically rejects utterances during training that do not sufficiently agree with previous utterances. Thus, from time to time the user may be reprompted for a word because the previous utterance was rejected. This is caused by the VRM hardware function and not the software. The user is prompted to enter the number of the first and last words to be trained in the vocabulary. Any contiguous subset of words may be selected. The user is then prompted to enter the number of training passes desired (0-64). If 0 training passes is chosen for a one-word vocabulary, the recognizer defaults to 5. For larger vocabularies, the recognizer uses the previous value for the number of training passes. If no previous training was conducted, the software reprompts the user for the number of training passes desired. Again the handling of a zero entry for the number of training passes is driven by the VRM hardware function. When training is complete, a message appears on the terminal, and the user is returned to the main menu.

Update. The update option is basically the same as train except that the on-board memory of the recognizer is not initialized. The user interface is identical. Reference patterns are updated by averaging in additional input from the speaker. On the other hand, the train procedure develops entirely new patterns. When recognition problems occur, the user must decide whether training new patterns or updating old patterns would be more beneficial. However, if a few update passes do not improve recognition, retraining is in order. If recognition is still poor after retraining, the vocabulary or the recognizer parameters should probably be changed.

Recognize. Recognize causes the VRM to compare incoming spoken words to stored templates. The vocabulary items being tested are randomly prompted one time each; the user speaks each word; and the software responds with the word that was recognized, that no word was recognized, that the utterance was too long, or that the word must be repeated.

Obviously, for the comparison required during recognition, word patterns must reside in the on-board memory of the recognizer. The following procedures cause patterns to be stored in RAM: TRAIN, UPDATE, GET WORD PATTERNS.

Two statistics files (stats.dat and stats.ana) are automatically generated by the recognize procedure. Each file contains the following data on each word prompted:

- (1) prompted word,
- (2) winning word,
- (3) winning score,
- (4) difference in delta scores, and
- (5) runner-up word of top two words.

The stats.dat file includes summary statistics of the difference score and reject threshold used, the number of words prompted, the number of words recognized, the number of words rejected, the number of utterances that were too long, and the number of words where the difference between the delta scores of the top two words was less than the difference-score criterion. A suggested reject-threshold (RTHL) is calculated based upon the average delta score of each of the winning words when the utterances were rejected. The stats.ana file contains similar information in a format readable

by the analyze program. More information on the use of the stats.ana file is provided in the section on data analysis.

File word patterns. The file-word-patterns option causes patterns residing in the on-board memory of the recognizer to be written to a VAX disk file. The software prompts the user to enter the name of the file where the patterns are to be stored (maximum of 63 characters). For consistency in file naming, all pattern files should be given a file type of PAT. The file name should be selected so that critical information such as the speaker's name, the vocabulary used, and the number of training passes is evident. Pattern files should reside in the user's own VAX directory, not in the ONR directory. The user must then indicate whether the file to be created is new or old. In addition, the user selects the first and last number of the vocabulary patterns to be stored. Any contiguous subset of pattern stored in the recognizer can be selected. The following data are requested:

- (1) file name for pattern storage,
- (2) status of pattern file (new or old),
- (3) number of first vocabulary word to be stored,
- (4) number of last vocabulary word to be stored,
- (5) speaker's name, and
- (6) complete file name of the vocabulary prompts.

In addition to this information the current values of T1, T2, MINSM, ETHL, and reject-threshold are stored. The number of training passes and update passes for the reference patterns are also stored. Filing the reference patterns is the only way to save them. When the recognizer is turned off, the patterns are lost unless they have been stored in a VAX disk file.

Get word patterns. The get-word-patterns option retrieves a pattern file from the VAX and puts the patterns in the VRM RAM. The user is prompted to enter the complete name of the VAX file where the patterns are stored, eg., disk-drive:[username]words.pat. (Files of reference patterns are created by selecting the file-word-patterns option.) The user then provides the numbers of the first and last pattern to be downloaded. After the patterns have been loaded into the VRM on-board memory, information concerning the reference patterns is displayed on the terminal screen. This information was stored when the file was created (see FILE WORD PATTERNS). The following information is displayed:

- (1) speaker's name (reference pattern speaker),
- (2) file creation date, and
- (3) name of the vocabulary used.

The current parameter settings of the recognizer and those when the file was created are displayed. These include T1, T2, ETHL, MINSM, and RTHL.

Of all the user options, get-word patterns and file-word patterns take the longest to run. Retrieving the maximum number of reference patterns (100) takes more than 5 sec. The user is notified when the get-word-patterns option is finished.

Recognition Problems

If the VIS is not processing utterances, it is probably because of one or more of the following:

- (1) the microphone is turned off,
- (2) the audio-level switch is off,
- (3) the audio-level switch is set too low,
- (4) the user is speaking too softly,
- (5) the microphone is positioned incorrectly, or
- (6) parameters are set incorrectly.

If the VIS is failing to recognize many utterances (reject errors), possible solutions are:

- (1) retraining reference pattern(s),
- (2) updating reference pattern(s),
- (3) decreasing the reject-threshold,
- (4) decreasing the difference-score criterion,
- (5) changing parameters, or
- (6) pausing longer between utterances.

If the VIS is recognizing words incorrectly (substitution errors) possible solutions are:

- (1) retraining reference pattern(s),
- (2) updating reference pattern(s),
- (3) increasing the reject-threshold,
- (4) increasing the difference-score criterion,
- (5) changing the vocabulary, or
- (6) changing the parameters.

If the VIS is recognizing words that are not included in the "legal" vocabulary (false accept errors), possible solutions include:

- (1) retraining reference pattern(s).
- (2) increasing the reject-threshold, or
- (3) changing the vocabulary.

TASK ENVIRONMENTS

One goal of the original software implementation of the VIS was to provide a package of software modules that could be modified and variously connected to develop a variety of task environments for experimentation. The software modules were initially organized into a menu-driven environment that implemented the host functions of the VRM in a user-friendly manner. The modules have since been reorganized into several additional types of environments. These are a simple prompt-and-recognition task, a maze task, data entry by form-filling, and the GENIE environment (Lindquist, Fainter, Guy, Hakkinen, and Maynard, 1982). One method of distinguishing among these environments is by the degree of user control of the dialogue. In the simple word recognition environment where the user is prompted to say words selected randomly from a vocabulary list, the computer controls the course of events completely. In the maze and form-filling environments, control is shared between the user and the computer. In the GENIE environment, which uses a command language, the user is in complete control of the sequence of the dialogue with only the syntax constraints of the command language.

The following sections explain the organization of these modules, the modifications that were made in order to develop new task environments, and a discussion of problems encountered when adding feedback and error-correction alternatives.

Software Modifications Required

Because the original software was written to provide for rapid modification, very few changes were necessary to adapt the software modules

implementing the VRM functions to different uses. The first change involved the manner in which data required to set up and operate the VIS were obtained. In the original menu-driven software, the user is prompted to provide necessary data. For the task environments, a new module was written so that all the data necessary for the VIS operation could be read from a file. The data read from this file are the values for:

- T1
- T2
- MINSM
- ETHL
- Reject Threshold
- Difference-Score Criterion
- Vocabulary Size
- Index Number of First Vocabulary Word
- Index Number of Last Vocabulary Word

and the names of the following data files:

- Vocabulary File
- Pattern File
- Results File
- Data-Analysis File.

The prompts to the user in the various modules were simply replaced by read statements (to read from this new experimenter's file) grouped together in this new module.

So that the experimenter's file could be developed easily and correctly by any user, a program was written to assemble and store the required information automatically. With this program, the necessary information is gathered from the pattern file (named by the experimenter) and from information provided by the experimenter. All information is checked for being within the proper bounds, and the existence of the vocabulary file and pattern file are confirmed.

With the incorporation of these changes, new environments can be easily developed by making changes only to the RECOGNIZE procedure which deals with determining the word spoken and the SCREEN procedure which formats the user's display and writes output to the user.

Feedback and Error-Correction Alternatives

Feedback. Because recognition errors do occur, user feedback is an important, but minimally explored, area of dialogue for voice input. Therefore, the alternatives for the type and mode of presentation of feedback were given much consideration. Feedback alternatives possible range from none to acknowledging that an utterance was received to displaying the recognized word or phrase and allowing the user to approve or disapprove it. Both visual and auditory modes of feedback were considered.

The full range of feedback alternatives has been implemented so that the type of feedback can vary depending upon the objectives of the specific environment or experiment. To change the type of user feedback, the calls to the SCREEN procedure within the RECOGNIZE procedure must be altered. This was accomplished by including the feedback alternatives in the SCREEN procedure and using data in the experimenter's file to determine which feedback would be available. In summary, the following types of feedback are available, and any combination can be selected by the experimenter.

No Feedback

Category Feedback (word recognized or not)
Auditory (tone)
Visual (light)

Word-by-Word Feedback
Auditory Shadowing (synthesized speech)
Visual Shadowing (display terminal)

Field/Command Summary Feedback:
Auditory (synthesized speech)
Visual (display terminal)

Specific auditory feedback can be provided either by calling stored vocabulary from the Votrax ML-1 Synthesizer in which all vocabularies words must be individually programmed and stored or by sending character strings to the Votrax Type 'N Talk which incorporates a text-to-speech conversion algorithm. Although the Type 'N Talk provides the easiest means of auditory feedback, word intelligibility is considerably better with the Votrax ML-1.

The experimenter must also decide whether or not to allow prompting and user confirmation of the first-choice word (and the second-choice word, if necessary) in cases where the difference between the delta scores of the top two words does not exceed the difference criterion established by the experimenter. With a vocabulary containing words that are similar phonetically, user prompting and confirmation of the first- and second-choice words might be highly desirable.

Error correction. The functionality of various types of feedback is intertwined with the type of error correction available to the user. The programmer, experimenter, and dialogue designer have a great deal of flexibility in selecting the type of error correction to provide when voice input is being incorporated into new software. However, providing error correction when voice input is added to an existing system depends to a great extent upon the degree to which the existing software can be altered.

In the prompt-and-recognition environment, either no error correction is provided or the user is permitted to confirm or deny the first- and second-choice words when the difference-score criterion is not met.

In the form-filling tasks two approaches to error correction were possible. Either the user could be required to confirm the correctness of every word or phrase recognized by the hardware, or the user could be

allowed to reenter words or phrases incorrectly recognized once an error was perceived by the user. Based upon efficiency considerations, a decision was made to provide error correction rather than require constant word or field confirmation. If the user were required to confirm every utterance, the rate of data entry would have been reduced by at least one-half with a related decrease in user satisfaction with the voice-entry dialogue. Two levels of error correction were provided in the form-filling task, each requiring a one-word command. One command is used to cancel the entire current field of data and another is used to delete only the most recent word uttered. By using these one-word commands, the user can correct any type of recognition error. The experimenter may select no error correction, last-word error correction, current-field error correction, or a combination of last-word and current-field correction.

The problem of providing error correction when voice input is used in the GENIE environment is probably more typical of the real-world situation where voice input is added to an existing system driven by a complex software package. In many cases the software of existing systems cannot be altered. Therefore, voice entry must look identical to keyboard entry for the software. In the case of GENIE, command processing occurs whenever a "carriage return" is detected. To allow the user to correct single-word voice entries, it was necessary to build an input buffer between user input and the language processor. The contents of the input buffer are sent to the language processor only when the voice equivalent of a keyboard "carriage return" is received. Thus, individual words in the entry can be corrected prior to a command terminator without requiring cancellation and re-entry of the entire command. By requiring an explicit terminator, user input by voice

parallels user input by keyboard. Any term may be selected as the equivalent of the "carriage return." Both single-word correction and command cancellation are provided.

However, the need for a command terminator for voice entry is somewhat awkward and requires additional data entry by the user and processing time by the hardware. If the programmer/dialogue author has access to the system software, the language processor could be rewritten to process commands word-by-word thereby avoiding the need for a command terminator. This alternative has also been provided in the GENIE environment. The only drawback is that because command processing occurs immediately upon recognition of a word, no single-word error correction is possible. Only command cancellation is available.

The most difficult, but probably the best, approach would be to provide software to act as a preprocessor that would recognize the end of a command through a set of syntax rules. When a valid command is completed, it is automatically sent to the language processor without the need for an explicit terminator. In addition, because input is buffered, error correction of single words or entire commands is possible. This implementation is not currently available for GENIE.

Experimental Task Environments

Prompt/recognition. In this environment the user is prompted to say words randomly selected from the vocabulary list. Three different implementations of this task are available and are distinguished by the feedback provided to the speaker. In the first task a word is selected and displayed, the speaker's utterance is processed, and a message is written to the terminal to indicate that an utterance was received. This procedure is

repeated until all the words have been prompted the desired number of times. This task, which is identical to the procedure provided to test recognition in the software for pattern development, can be used to test the recognition accuracy of the VIS using various settings of the hardware parameters, different vocabularies, or different numbers of training passes used to establish reference patterns.

In another simple recognition task two bar graphs are displayed after each word is recognized. The heights of the bar graphs represent the value for the winner's score and the difference between the delta scores of the top two words. The only software change was to alter the SCREEN procedure to display the bar graphs after each recognition. This task can be used to demonstrate to new users how much recognition accuracy (as measured by reject threshold and the difference score) can vary if user input is not consistent.

Maze. In the maze task users enter voice commands to move through a maze. Feedback is provided both by the movement of the cursor through the maze and by display of the recognized word. This simple task can be used to demonstrate discrete-word recognition or to familiarize new users with voice input. The only software change required was to alter the SCREEN procedure to change the maze after each recognized word.

Form-filling. In the form-filling task, a form consisting of seven data areas is displayed. The user's task is to transcribe the data necessary to fill the form. A new form is presented to the user either when the user has asked for the previous form to be filed or when a specified period of time has elapsed. By selecting the particular method by which the software determines when to display a new form, the experimenter may test voice input in either a

time-driven or event-driven environment. In the time-driven environment the pacing of the task is under computer control, whereas in the event-driven environment the user determines the speed of presentation of the new forms. In both cases, the types of feedback available for experimental manipulation are identical. If the delta score obtained does not exceed the difference-score criterion established by the experimenter, the speaker may be prompted to confirm the first-choice (and if necessary the second-choice) word. If this option is desired, the experimenter selects confirmation of "not sures" when queried by the program that creates the experimenter's file. In addition, the experimenter may select the specificity and sensory mode of feedback. The specificity of feedback can range from none to categorical (word recognized or not) to word-by-word shadowing (exact word recognized) to summary feedback at the end of data entry for a field. The mode of feedback may be visual (light or visual display), auditory (tone or synthesized speech), or both.

In both form-filling tasks the necessary changes to the software were minimal. The SCREEN procedure was used to write the form on the terminal display. The procedure had to be modified to display a form rather than the split-screen format used in the basic prompt-and-recognition environment. In the user-paced version, the RECOGNIZE procedure is used to determine when to write a new form. In the computer-paced version, the RECOGNIZE procedure was altered to call the SCREEN procedure for a new form after a specific time period has elapsed. In addition, a new procedure called PREPWORD was written to determine the type of word recognized and to set the cursor parameter used in the SCREEN procedure so that the recognized word or field is displayed in the proper location on the screen.

GENIE. The third type of environment provided to test voice input was GENIE (Generalized Task for Interactive Experiments) in which a command language is used to control the task. This environment runs under the Dialogue Management System (DMS) described by Ehrich (1982). The voice-interface software for GENIE configures the VRM, downloads the pattern file, and puts the recognizer in recognition mode to wait for input from the VIS. If a spoken word is recognized, the word is sent to the GENIE language processor or stored in the input buffer, depending upon whether a command terminator is necessary to begin command processing. In addition, the recognition data are used to provide some form of auditory or visual feedback to the user. When a command is completed (either by the command terminator or the language processor receiving enough input), the appropriate action is taken by the GENIE software. The only software changes necessary to implement voice entry for GENIE were to rewrite the RECOGNITION procedure so that calls to the SCREEN procedure were replaced with REQUESTs (a DMS command) to the GENIE process. The SCREEN procedure was not used because GENIE incorporates its own display formatter. To provide a means for error correction by cancelling the entire command, the GENIE language processor was rewritten to recognize the word "cancel." In addition, the PASCAL functions writeln and readln were replaced with procedures to handle input and output from the VIS. Numerical codes for the index numbers of the recognized words had to be translated to the associated character string prior to being sent to the GENIE language processor.

DATA ANALYSIS

An interactive software tool for analyzing vocabularies for voice input has been developed because these analyses are critical, but time-consuming. This tool is to aid the experimenter and/or dialogue author who needs to examine the recognition problems with a given vocabulary. These problems include, but are not limited to, word rejections and substitution errors. The tool can automatically produce tables of word confusions, winner's scores, and difference scores. The data upon which this tool acts are collected from programs that were designed to train people to use voice recognition equipment or to complete a specific task using voice input. Examples of the types of experimental tasks have already been presented.

This system has been designed to be menu-driven, but may also be command-driven. Once an experimenter becomes familiar with the available commands, that person may work more rapidly by remaining at the command level instead of returning to the menu each time a command is completed.

These data analysis procedures could also be used dynamically to analyze recognition rates as the equipment is being used provided a means is established to enter the actual words spoken by the user. Possible applications include monitoring between-word confusions to determine whether a change in vocabulary is desirable. This capability could be particularly useful in applications where users select their own vocabulary for system functions, and the system design has no control over substitution errors caused by similar-sounding vocabulary items. With on-line data analysis, confusion problems could be readily detected, the user could be notified of the problem and asked to select a new term for one item in the confused pair, and pattern training for the new item could be quickly completed. Monitoring

and analyzing recognition accuracy during recognizer usage may also be valuable for detecting recognition errors caused by fatigue, stress, or other voice changes. Whenever recognition drops below a specified criterion, the vocabulary items involved can be retrained immediately.

Software

The entire system is written in DEC PASCAL. PASCAL was chosen for several reasons. The data structures (RECORD types) available within the language are well fitted for the task (each node of the tree that contains the data is a RECORD) and the built-in NEW storage-allocation function made the task of creating a variable-sized tree trivial. In addition, each command action corresponds to a value of an enumerated type. If a new command is desired, this type is extended and the procedure to do the necessary computation is added to the analysis module.

The software procedures to process the data produced in the various studies are divided into two primary parts, one to control the analysis and another to do the actual analysis. In addition to drawing the menu and receiving user commands, the control software is also responsible for opening and closing the files used for input and output and obtaining the raw data.

Before the menu is drawn, the experimenter or dialogue author is queried for the name of the file in which the data to be analyzed resides. The existence of this file is checked. If the file does not exist, the user is asked either to re-enter the file name or to enter the word "quit" to exit from the system. If the input file exists, it is opened and the information is read according to one of two formats. For both formats, the initial elements are the same. They are:

- (1) whether the user utterances were prompted,
- (2) speaker's name,
- (3) speaker's sex,
- (4) vocabulary used,
- (5) parameters used during training, and
- (6) parameters used during the study.

The recognition data may be completed in one of two ways depending upon whether the spoken words were prompted. If the study involved word prompting, then the data can be completed automatically because the system knows both the word recognized and the word spoken (provided the speaker says the prompted word). If the study involved a data entry or control task, where the word actually spoken is not known by the system, as in the GENIE environment, then each word spoken must be entered by the experimenter after the trial is completed. This involves inserting the word spoken before the recognition data for that utterance. This may be accomplished through the use of a text editor or by using a program which prompts for the spoken word and inserts it into the file. In order to save the necessary data, a recording of the trial has to be made. It is recommended that a recording also be used for studies in which the words are prompted in order to insure that the subject has spoken the prompted word. If there are any instances where the spoken word and the prompted word are not the same, they should be removed from the results file.

The first five elements of the recognition data are the same for either type of study. They are:

- (1) prompted (spoken) word,
- (2) winning word,
- (3) runner-up word,
- (4) winner's score, and
- (5) delta score.

If user input was prompted and the recognition data were completed automatically, there are two additional values available: the position numbers in the vocabulary of both the prompted and winning words. If these values are not present, they are computed as the data are read.

In either case, a tertiary tree is constructed from the data. Each node corresponds to one utterance and contains all of the information outlined above. The order of the tree is based on the prompted (spoken) word. (A shortened and constructed example of the tree is given in Figure 1.)

After the tree data are read and the tree constructed, the user is asked for the name of the file to which the results of the analysis are to be written (output file). This file is opened, and the header information from the data file is written to it. The data analysis may now begin.

The menu presented to the user is given in Table 7. All user selections are validated. If the command is found to be invalid, the user will be prompted to enter a correct command. The menu remains on the screen until a valid command is entered. The menu is then erased and replaced by queries for information needed to complete the command or by the results produced by the actions of the command. When the command is completed, the user is asked to either enter another command or to enter a null line to return to the main menu.

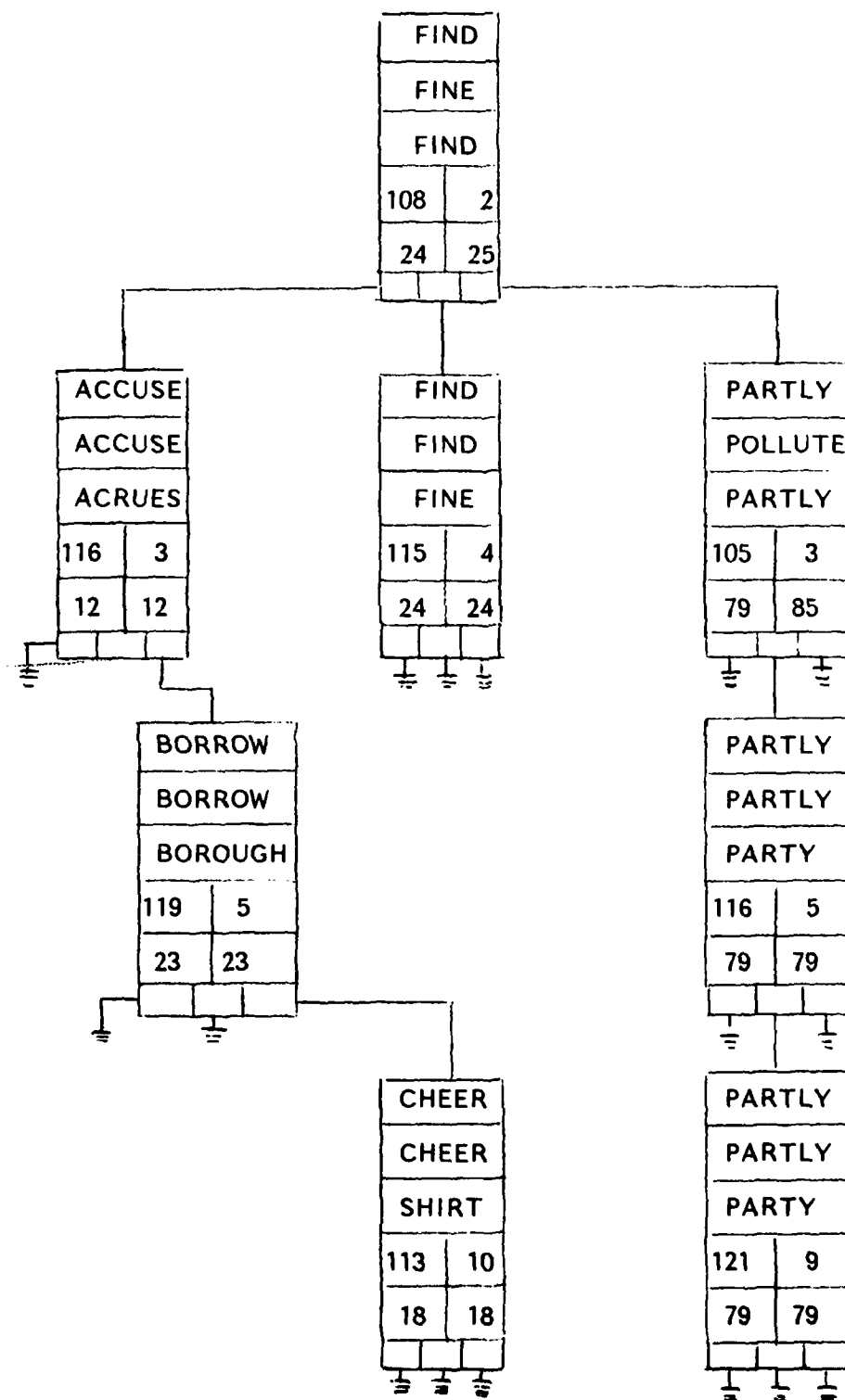


Figure 1. Example of a tertiary tree used in data analysis.

TABLE 7

Main Menu for the Data-Analysis Program

ANALYSIS COMMANDS

- (a)lphabetic summary
- (c)onfusion problem words
- (d)elta score table
- (exit)
- (i)nformation loss
- (l)ist delta < specified value
- (m)isrecognitions -- substitution errors
- (more) data files
- (p)rompted vs. recognized table
- (s)tatistics
- (r)ecognized for RTHL and DELTA

Please enter the desired command. _____

Each command, with the exception of EXIT and MORE, corresponds to an action passed to the analysis routine. The actions performed by each command are described below. For most of the commands, a traversal of the tree is made to find all occurrences of events that correspond to the desired action. An example of this is the alphabetic summary of words. A modified inorder traversal of the tree is performed. When the left pointer of a node is nil, the appropriate information is printed. Then a traversal is made of the repetitions of the prompted (spoken) word. Next, a traversal of the right side of the subtree is made. The commands "delta," "prompted," and "statistics" do not require traversals, because these results are computed as the data are placed into the tree. The commands "alphabetic," "delta," "misrecognitions," "prompted," and "statistics" may only be used once per data file because all possible output for the command is obtained when the command is first used. An attempt to issue any of these commands again will result in a warning message and the menu being redrawn.

All data created by the analysis program are written to an output file. Several of the commands also produce immediate output to the display screen. However, most of the commands only write output to the output file. The contents of this disk file can be printed in hardcopy, if the user desires.

Command Functions

The following sections give brief explanations of the results of the various analysis commands. Samples of the output from these procedures are given in Appendix B.

Alphabetic summary. The results of this command are not displayed on the screen but are sent directly to an output file. A line is written on the display to inform the user that the command is being processed and another

is written when the command has been completed. An alphabetic summary of the prompted or spoken words is produced. A list of all words that were recognized from subject utterances is given. This list includes the winning word, the runner-up word, the winner's score, and the delta score. If the winning word is not equal to the word spoken, either because of a substitution error or non-recognition, the line which contains this information is flagged by being preceded by a series of asterisks. Two additional summaries are produced including either utterances where the recognized word was the same as the word spoken (prompted) or substitution errors. At the end of each summary, the mean and standard deviation for each parameter are given. At the end of the summary, the number of utterances in the file, the number of words not recognized, the number of substitution errors, and the number of utterances that were too long are listed.

Confusion problems. This command provides a rapid check for confusion problems associated with a specific word. Output is displayed on the screen as well as written to the output file. The user is asked for a word from the vocabulary to be checked and a difference-score criterion value to be used. All entries are searched for this word, both as the winning word and as the runner-up word. For all occurrences found where the difference between the winning and runner-up words is less than the difference criterion specified, both words are written to the screen in the format:

"word1" would not have been distinguished from
"word2" at difference score = xx.

If there are no entries which meet the specifications, a message to that effect is displayed. The user is then asked to enter another word and difference

criterion or to type "quit" to return to the command level. This command may be reissued during the analysis session.

Delta-score table. A table is produced depicting the delta values (difference between the winner and runner-up words) for all the words spoken during the trial. This table is not written to the display screen, but to an output file.

Exit. The exit command causes the user to exit from the analysis program. Before exiting, the user is asked if the output file is to be saved. If it is not, then it will be deleted automatically. If the user wants to print a copy of the output file, it must be saved.

Information loss. Values are computed for the entropy, equivocation, and relative information loss for the vocabulary. The formulas used are those put forward by Woodard and Nelson (1982).

List delta \leq specified value. The user is queried for a value for the difference-score criterion. This value is used to compile a list of all entries which have a delta value less than the criterion specified. This list is written to the screen as well as to the output file.

Misrecognitions. If there have been any false accepts or substitution errors during the trial, they will be written to the screen. If there have been none, a message to that effect will be displayed.

More data. This command is entered when the user wants to examine more than one data file. The actions of this command are: (1) ask if the output file being used for the analysis in progress is to be saved; (2) close the output file, and (3) if desired, delete the output file. Then the user is asked to enter the name of a new data to be analyzed and an output file.

Prompted vs. recognized table. A confusion matrix of the vocabulary will be produced. The horizontal axis of the table is the word number of the recognized utterance. The vertical axis is the word number and word which was spoken (prompted). This table is not displayed but is sent to the output file. For each vocabulary item, a total is kept of the word(s) recognized along with the total number of times it was spoken (prompted). For each vocabulary item, this information is written underneath its proper heading. In each table, the entire vocabulary is listed with the frequency of recognition of 30 items. If there are more than 30 words in the vocabulary, the table is reproduced with the results for the next 30 words. This process is repeated until the confusion matrix for the entire vocabulary is depicted. This table may be used to see confusion problems within the vocabulary rapidly.

Statistics. This command produces several types of summary statistics. The name of the speaker along with the values of all training parameters are given. Then the parameter values used during the trial are presented. The mean and variance of the winner's score and the delta value are computed for each word, for all recognized words that were the same as the prompted (spoken) word, for all substitution errors, and for the entire vocabulary. The minimum and maximum values of the winner's score and delta for each word and for the overall vocabulary are listed. Suggested values of the reject threshold and difference-score criterion are given. Finally, values are given for the ratio of the delta score to the winner's score, by word and by vocabulary.

Recognized for RTHL and DELTA. The user is prompted for values of the reject threshold and difference-score criterion. These values are then

used to calculate the number of utterances that would have been recognized from the total sample. The results are written to the display screen as well as to the output file.

REFERENCES

- Ehrich, R. W. The DMS Multiprocess Execution Environment. Blacksburg, VA: Virginia Polytechnic Institute and State University, Department of Computer Science, Technical Report CSIE-82-6, April, 1982.
- Lindquist, T. E., Fainter, R. G., Guy, S. R., Hakkinen, M. T., and Maynard, J. F. GENIE: A Dialogue-Rich Task Environment for Empirical Studies of Human-Computer Interactions. Blacksburg, VA: Virginia Polytechnic Institute and State University, Department of Computer Science, Draft Report, August, 1982.
- Interstate Electronics Corporation. Voice Recognition Module (VRM) Reference Manual (rev. 3). Anaheim, CA: Author, April, 1981.
- Nye, M. Human factors analysis of speech recognition systems. Speech Technology, 1982, (2), 50-57. (a)
- Nye, M. Voice integration - The critical mass. Proceedings of the Voice Data Entry Systems Applications Conference. Sunnyvale, California: Lockheed Missiles and Space Company, 1982. (b)
- Woodard, J. P. and Nelson, J. T. An information theoretic measure of speech recognition performance. Proceedings of the Workshop on Standardization for Speech I/O Technology. Gaithersburg, Maryland: National Bureau of Standards, March, 1982.

APPENDIX A

VRM COMMAND SUMMARY

Key:

- 1 = Host-to-VRM Command
- 2 = VRM-to-Host Intermediate Message
- 3 = Host-to-VRM Intermediate Message
- 4 = VRM-to-Host Acknowledge Message

SETCHARS

Purpose: Change framing characters, acknowledge character, nonacknowledge character, or termination character; required after power up and after hardware or software reset.

- 1. STX,!,&,-,¢,%,#,CR,CR
- 2. None
- 3. None
- 4. &,X,CR

RESET

Purpose: Cause VRM to set all flags to zero, to set all word boundaries to default values, and to read all hardware settings; does not alter current reference pattern

- 1. !,3,CR
- 2. None
- 3. None
- 4. R,3,CR

SETFLAGS

Purpose: Set or clear mode flags.

1. !,A,1,1,0,0,1,1,CR
2. None
3. None
4. &,A,CR

SETPARMS

Purpose: Write values for word-boundary parameters.

1. !,C,T1,T2,ETHL,MINSM,CR
2. None
3. None
4. &,C,CR

Purpose: Read values of word-boundary parameters.

1. !,D,CR
2. None
3. None
4. &,D,T1,T2,ETHL,MINSM,CR

SET REJECT THRESHOLD

Purpose: Set value for reject threshold.

1. !,4,zzz,CR
2. None
3. None
4. &,4,CR

TRAIN

Purpose: Initialize VRM reference-pattern area, provide indices, and generate patterns for each vocabulary item.

1. !,1,xx,yy,bb,CR
2. &,8,uu,CR
3. None
4. &,1,CR

UPDATE

Purpose: Provide indices and generate reference patterns for each vocabulary item; VRM reference-pattern area is not initialized.

1. !,2,xx,yy,bb,CR
2. &,8,uu,CR
3. None
4. &,2,CR

RECOGNIZE

Purpose: Cause VRM to compare incoming words with the reference patterns specified; the index number of the word with the highest score will be output.

1. !,9,xx,yy,CR
2. &,9,ww,dd,sss,rr,CR
3. None
4. &,z,CR

FILE-WORD-PATTERNS

Purpose: Request VRM to transmit (upload) reference patterns of specified class to host for storage.

1. !,7,xx,yy,CR
2. €,Reference Pattern Data,CR
3. None
4. None

GET-WORD-PATTERNS

Purpose: Transmit (download) to the VRM from the host reference patterns of the specified class. Replace part or all of the previous contents of VRM memory.

1. !,6,xx,yy,CR
2. &,6,CR
3. ~,Reference Pattern Data,CR
4. %,CR or #,CR

<u>Symbol</u>	<u>Definition</u>
STX	Control Character to Initiate Command to Change Control Characters
!	Framing Character 1 (default = DC1)
&	Framing Character 2 (default = DC2)
~	Framing Character 3 (default = DC3)
©	Framing Character 4 (default = DC4)
%	Acknowledge Character (default = ACK)
#	Nonacknowledge Character (default = NAK)
CR	Terminator Character of Carriage Return
X	VRM-to-Host Acknowledge Character to Change Control Characters
3	Command Identifier for Reset
A	Command Identifier for Set Flags
C	Command Identifier for Write Word-Boundaries
T1	Beginning of Word Threshold (2-digit number between 16 and 64)
T2	Continuing Speech Threshold (2-digit number between 08 and 64)
ETHL	Maximum Number of Nonsignificant Samples Allowed During Utterance (2-digit number between 08-64)
MINSM	Minimum Number of Significant Samples for Sound to be Processed (2-digit number between 16 and 32)
D	Command Identifier for Read Word-Boundaries
4	Command Identifier to Set Reject Threshold
zzz	Reject Threshold (3-digit number between 000 and 128)
1	Command Identifier for Train
xx	Item Number of First Vocabulary Word (2-digit number between 00 and 99)

yy	Item Number of Last Vocabulary Word (2-digit number between xx and 99)
bb	Number of Training Passes (2-digit number between 01 and 64)
8	Command Identifier for Train
uu	Vocabulary-Item Number for Prompt (2-digit number between 00 and 99)
2	Command Identifier for Update
Z	Indicator to Continue Other Processing until VRM Receives Spoken Input
9	Command Identifier for Recognize
ww	Winning-Word Index (2-digit number between 00 and 99), FF if none exceeded RTHL, or LL if utterance exceeded 250 significant sample
dd	Last Two Digits of Difference Between Winner and Runner-up Scores
sss	Winner's Score (3-digit number between 000 and 128)
rr	Runner-up Word Index (2-digit number between 00 and 99)
7	Command Identifier for Upload
6	Command Identifier for Download

APPENDIX B

Sample Output from Data-Analysis Routines

PROMPTED WORD	WINNING WORD	RUNNER-UP WORD	WINNERS SCORE	DELTA
EIGHT	EIGHT EIGHT EIGHT EIGHT EIGHT EIGHT EIGHT EIGHT EIGHT EIGHT	TWO SIX TWO SIX TWO SIX SIX SIX SIX SIX	121 116 119 119 119 119 119 119 119 119	22 19 19 20 20 20 21 21 21 21
FIVE	FIVE FIVE FIVE FIVE FIVE FIVE FIVE FIVE FIVE FIVE	NINE NINE ONE NINE ONE NINE NINE NINE NINE NINE	108 115 114 118 118 125 125 114 123 110	13 17 20 26 26 23 23 16 18 21
FOUR	FOUR FOUR FOUR FOUR FOUR FOUR FOUR FOUR FOUR FOUR	ONE ONE ONE ONE ONE ONE ONE ONE ONE ONE	119 125 125 118 125 125 125 125 125 119	8 17 24 17 24 24 24 24 24 28
NINE	NINE THREE THREE NINE NINE NINE NINE NINE NINE NINE	THREE NINE THREE NINE THREE THREE THREE THREE THREE THREE	106 108 107 107 101 101 119 108 105 103	6 3 3 3 8 8 16 12 5 3

ONE	ONE ONE ONE ONE ONE ONE ONE ONE ONE ONE	FOUR FOUR FOUR FOUR FOUR FOUR FOUR FOUR FOUR FOUR	119 120 118 122 115 119 119 122 116 116	20 16 14 14 18 18 20 24 17 17
SEVEN	SEVEN SEVEN SEVEN	ONE SIX SIX	113 119 119	12 12 11

**** ALPHABETIC SUMMARY OF CORRECT RECOGNITIONS ****

WINNING WORD	RUNNER-UP WORD	WINNERS SCORE	DELTA
EIGHT	TWO	121	22
EIGHT	SIX	116	19
EIGHT	SIX	113	16
EIGHT	TWO	115	21
EIGHT	SIX	119	17
EIGHT	SIX	121	22
EIGHT	TWO	119	20
EIGHT	SIX	116	17
FIVE	NINE	108	17
FIVE	NINE	110	17
FIVE	NINE	112	18
FIVE	NINE	117	19
FIVE	NINE	117	19
FIVE	NINE	117	19
FIVE	NINE	118	20
FIVE	NINE	118	20
FIVE	NINE	119	21
FIVE	NINE	115	17
FOUR	ONE	119	8
FOUR	ONE	119	8
FOUR	ONE	125	23
FOUR	ONE	128	24
FOUR	ONE	135	25
FOUR	ONE	118	17
FOUR	ONE	125	16
FOUR	ONE	111	27
FOUR	ONE	106	4
NINE	THREE	103	3
NINE	FIVE	108	12
NINE	THREE	111	18
NINE	FIVE	114	19
NINE	THREE	107	8
ONE	FOUR	119	20
ONE	FIVE	119	17
ONE	FOUR	112	24
ONE	FOUR	119	20
ONE	FOUR	115	18
ONE	FOUR	112	24
ONE	FOUR	118	19
ONE	FOUR	120	16
SEVEN	ONE	123	18
SEVEN	SIX	119	19
SEVEN	SIX	115	16
SEVEN	SIX	120	16
SEVEN	SIX	116	17
SEVEN	SIX	122	23

**** ALPHABETIC SUMMARY OF MIS-RECOGNITIONS ****

PROMPTED WORD	WINNING WORD	RUNNER-UP WORD	WINNERS SCORE	DELTA
NINE	THREE	NINE	105	1
NINE	THREE	NINE	103	3
NINE	THREE	NINE	98	1

For the mis-recognitions:

The mean winning score was: 101.33

The standard deviation of the winning score was: 3.512

The mean difference score was: 3.00

The standard deviation of the difference score was: 2.000

the winners score variance = 5.597
the delta score mean = 17.90
the delta score variance = 5.608

RELATIVE INFORMATION LOSS

ENTROPY: 3.3207

EQUIVOCATION: 0.1034

RIL: 0.0311

*** LIST OF WORDS WHICH HAVE DIFFERENCES LESS THAN ASKED FOR. ***

The prompted word was NINE	
The spoken word was NINE	
The runner up word was FIVE	7
The difference was	
The prompted word was NINE	
The spoken word was THREE	
The runner up word was NINE	3
The difference was	
The prompted word was NINE	
The spoken word was THREE	
The runner up word was NINE	1
The difference was	

*** LIST OF WORDS WITH CONFUSION PROBLEMS. ***

NINE
NINE
NINE

would not have been distinguished from FIVE
would not have been distinguished from THREE
would not have been distinguished from THREE

at delta = 4
at delta = 4
at delta = 4

WORD	MEAN / STANDARD DEVIATION TABLE					
	WINNERS SCORE		DELTA SCORE		WINNER / DELTA	(inclusive)
	MEAN	S. D.	MEAN	S. D.		
ZERO	119.44	2.698	16.22	3.114	7.34	
ONE	119.22	2.108	19.56	2.833	6.10	
TWO	118.70	2.710	16.40	3.718	7.24	
THREE	122.00	1.826	25.20	1.751	4.84	
FOUR	122.50	3.536	18.60	6.569	6.59	
FIVE	117.00	5.944	19.30	4.001	6.06	
SIX	117.70	3.093	18.00	4.216	6.54	
SEVEN	119.60	2.797	18.20	3.584	6.57	
EIGHT	118.00	2.828	19.40	2.171	6.08	
NINE	106.20	6.339	8.10	5.705	10.53	13.11

For the entire vocabulary

the winners score mean = 118.01

the winners score variance = 5.597

the delta score mean = 17.90

the delta score variance = 5.608

WORD	WINNERS		SCORE		DELTA SCORE	
	MIN	MAX	MIN	MAX	MIN	MAX
ZERO	115	123	9	19		
ONE	115	122	16	24		
TWO	115	123	10	23		
THREE	119	125	22	28		
FOUR	118	128	8	26		
FIVE	108	125	13	26		
SIX	111	122	11	25		
SEVEN	115	123	12	23		
EIGHT	113	121	16	22		
NINE	98	119	1	18		

For the vocabulary listed above values were:

Minimum winners score: 98
 Maximum winners score: 128
 Minimum delta score: 1
 Maximum delta score: 28
 Average winning score: 118
 Average delta score: 18

NUMBER RECOGNIZED FOR RTHL AND DELTA

For the reject threshold = 105 and delta = 8 the results are as follows:

total accepts: 92
correct accepts: 92
mis recognitions: 0

total rejects caused by reject threshold: 1
with first choice correct: 1
with second choice correct: 0

total rejects caused by delta value: 2
with first choice correct: 1
with second choice correct: 1

total unacceptable caused by reject threshold and delta value: 3
with first choice correct: 1
with second choice correct: 2

For the reject threshold = 112 and delta = 1 the results are as follows:

total accepts: 87
correct accepts: 87
mis recognitions: 0

total rejects caused by reject threshold: 11
with first choice correct: 8
with second choice correct: 3

total rejects caused by delta value: 0
with first choice correct: 0
with second choice correct: 0

total unacceptable caused by reject threshold and delta value: 0
with first choice correct: 0
with second choice correct: 0

For the reject threshold = 112 and delta = 8 the results are as follows:

total accepts: 87
correct accepts: 87
mis recognitions: 0

total rejects caused by reject threshold: 6
with first choice correct: 6
with second choice correct: 0

total rejects caused by delta value: 0
with first choice correct: 0
with second choice correct: 0

total unacceptable caused by reject threshold and delta value: 5
with first choice correct: 2
with second choice correct: 3

For the reject threshold = 112 and delta = 15 the results are as follows:

total accepts: 78
correct accepts: 78
mis recognitions: 0

total rejects caused by reject threshold: 1
with first choice correct: 1
with second choice correct: 0

total rejects caused by delta value: 9
with first choice correct: 9
with second choice correct: 0

total unacceptable caused by reject threshold and delta value: 10
with first choice correct: 7
with second choice correct: 3

For the reject threshold = 117 and delta = 3 the results are as follows:

total accepts: 71
correct accepts: 71
mis recognitions: 0

total rejects caused by reject threshold: 26
with first choice correct: 24
with second choice correct: 2

total rejects caused by delta value: 0
with first choice correct: 0
with second choice correct: 0

total unacceptable caused by reject threshold and delta value: 1
with first choice correct: 0
with second choice correct: 1

OFFICE OF NAVAL RESEARCH

Engineering Psychology Group

TECHNICAL REPORTS DISTRIBUTION LIST

OSD

CAPT Paul R. Chatelier
Office of the Deputy Under Secretary
of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Department of the Navy

Engineering Psychology Group
Office of Naval Research
Code 442 EP
Arlington, VA 22217

Manpower, Personnel & Training
Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Special Assistant for Marine Corps
Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Dr. J. Lester
ONR Detachment
495 Summer Street
Boston, MA 02210

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375

Department of the Navy

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D.C. 20375

Dr. J. S. Lawson
Naval Electronic Systems Command
NELEX-06T
Washington, D.C. 20360

Dr. Robert G. Smith
Office of the Chief of Naval
Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation
Group
Orlando, FL 32813

CDR Norman E. Lane
Code N-7A
Naval Training Equipment Center
Orlando, FL 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. Paul Heckman
Naval Ocean Systems Center
San Diego, CA 92152

Department of the Navy

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Human Factors Technology Administrator
Office of Naval Technology
Code MAT 0722
800 N. Quincy Street
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 334A
Washington, D.C. 20361

Commander
Naval Air Systems Command
Crew Station Design
NAVAIR 5313
Washington, D.C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 03416
Washington, D.C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 81323
Washington, D.C. 20360

Larry Olmstead
Naval Surface Weapons Center
NSWC/DL
Code N-32
Dahlgren, VA 22448

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab.
Naval Submarine Base
Groton, CT 06340

Dr. Robert Blanchard
Navy Personnel Research and
Development Center
Command and Support Systems
San Diego, CA 92152

Department of the Navy

CDR J. Funaro
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Dr. S. Schiflett
Human Factors Section
Systems Engineering Test
Directorate
U.S. Naval Air Test Center
Patuxent River, MD 20670

Department of the Army

Mr. J. Barber
HQs, Department of the Army
DAPE-MBR
Washington, D.C. 20310

Department of the Navy

Dr. Edgar M. Johnson
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and
Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U.S. Army Human Engineering Labs.
Aberdeen Proving Ground, MD 21005

Department of the Air Force

U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Department of the Air Force

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks Air Force Base, TX 78235

Other Government Agencies

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314

Other Organizations

Dr. Robert T. Hennessy
NAS - National Research Council (COHF)
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Dr. William R. Uttal
Institute for Social Research
University of Michigan
Ann Arbor, MI 48109

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Psychological Documents
ATTN: Dr. J. G. Darley
N 565 Elliott Hall
University of Minnesota
Minneapolis, MN 55455

Mr. Richard Main
ONR Resident Representative
George Washington University
2110 G. Street, N.W.
Washington, D.C. 20037

Dr. John J. O'Hare
Code 455
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Dr. E. Gloye
ONR Western Regional Office
1030 East Green Street
Pasadena, CA 91106

Other Organizations

CDR Norman E. Lane
Code N7A
Naval Training Equipment Center
Orlando, FL 32813

Dr. J. Hopson
HF Engineering Division
Naval Air Development Center
Warminster, PA 18974

Dr. A. Meyrowitz
Code 433
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Dr. Thomas McAndrew
Code 32
Naval Undersea Systems Center
New London, CT 06320

Mr. Walter P. Warner
Code KOZ
Strategic Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Mr. John Impagliazzo
Code 101
Newport Laboratory
Naval Underwater Systems Center
Newport, RI 02840

Dr. Mel C. Moy
Code 302
Naval Personnel R&D Center
San Diego, CA 92152

Dr. Richard Neetz
Pacific Missile Test Center
Code 1226
Pt. Mugu, CA 93042

Mr. Larry Olmstead
NSWC
Code N32
Dahlgren, VA 22448

Mr. Rick Miller
NSWC
Code N32
Dahlgren, VA 22448

Other Organizations

Dr. Arthur Fisk
ATT Long Lines
12th Floor
229 W. Seventh St.
Cincinnati, OH 45202

Dr. Lou Chmura
Code 7592
Naval Research Laboratory
Washington, DC 20375

Dr. R. J. K. Jacob
Code 7590
Naval Research Laboratory
Washington, DC 20375

ATE
MED
8